

WHAT IS CLAIMED IS:

1. A method of identifying one or more sequences of a target nucleic acid comprising:
  - a. contacting a target nucleic acid with a first set of pools of probes, wherein at least one pool in the set comprises a mixture of two or more probes having different sequences in information regions of the probes, under conditions which produce, on average, more probe:target hybridization with probes which are perfectly complementary to the target nucleic acid in the information region of the probes than with probes which are mismatched in the information regions;
  - b. detecting a first subset of pools for which a level of hybridization indicates that there is at least one perfectly complementary probe within each pool; and
  - c. identifying one or more sequences of the target nucleic acid from the first subset of pools detected in step (b) by compiling overlapping sequences of the information regions of the probes in the subset of detected pools, wherein one or more pooling false positive probes are eliminated as a result of compilation of overlapping sequences.
2. A method of identifying one or more sequences of a target nucleic acid comprising:
  - a. contacting a target nucleic acid with a first set of pools of probes, wherein at least one pool in the set comprises a mixture of two or more probes having different sequences in information regions of the probes, under conditions which produce, on average, more probe:target hybridization with probes which are perfectly complementary to the target nucleic acid in the information region of the probes than with probes which are mismatched in the information regions;
  - b. assigning a hybridization score to each probe wherein each probe within a pool is assigned the same hybridization score, and
  - c. identifying one or more sequences of the target nucleic acid by analysis of hybridization scores of overlapping probes, wherein one or more probes with false high scores arising from pooling of probes are eliminated by analysis of hybridization scores of overlapping probes.

3. The method of claim 2 wherein a statistical analysis of hybridization scores is performed in step (c).

4. The method of claim 3 wherein step (c) further comprises  
5 calculating a score for the identified one or more sequences of the target nucleic acid.

5. The method of claim 1 further comprising, following step (b) and before step (c), the steps of:

10 a. contacting the target nucleic acid with a second set of pools of probes containing at least one probe having the same information region as a probe in the first set,

b. detecting a second subset of pools for which the level of hybridization indicates that there is at least one perfectly complementary probe within each pool; and

15 c. eliminating probes with the same information regions present in both the first set of pools of probes and the second set of pools of probes that are not present in both the first detected subset of pools and the second detected subset of pools.

20 6. The method of claim 5 wherein the first and second sets of pools of probes comprise the same information regions.

7. The method of claim 5 wherein the first and second sets of pools of probes comprise the same probes.

25 8. The method of claim 2 further comprising, after step (b) and before step (c), the steps of:

a. contacting the target nucleic acid with a second set of pools of probes containing at least one probe having the same information region as a probe in the first set,

30 b. assigning a hybridization score to each probe wherein each probe within a pool is assigned the same hybridization score.

9. The method of claim 8 further comprising the step of:

c. eliminating the higher of two scores for probes present in both the first set and second set of pools of probes.

10. The method of claim 8 wherein the first and second sets of pools  
5 of probes comprise the same information regions.

11. The method of claim 8 wherein the first and second sets of pools of probes comprise the same probes.

10 12. The method of claim 1 or 2 in which the target nucleic acid is  
labeled.

13. The method of claim 1 or 2 in which the probes are labeled.

15 14. The method of claim 1 or 2 in which the label is a fluorophore.

15. The method of claim 1 or 2 in which the label is attached to a terminal nucleotide.

20 16. The method of claim 1 or 2 in which the label is attached to an internal nucleotide.

17. The method of claim 1 or 2 in which the first set of pools of probes is immobilized on one or more solid supports.

25 18. The method of claim 17 in which the pools of probes are arranged in a spatially-addressable array in which each pool has a unique address.

19. The method of claim 1 or 2 in which the target nucleic acid is  
30 immobilized on one or more solid supports.

20. A method of identifying one or more sequences of a target nucleic acid comprising:

0972600 - 040600

1000 999 998 997 996 995 994 993 992 991 990 989 988 987 986 985 984 983 982 981 980 979 978 977 976 975 974 973 972 971 970 969 968 967 966 965 964 963 962 961 960 959 958 957 956 955 954 953 952 951 950 949 948 947 946 945 944 943 942 941 940 939 938 937 936 935 934 933 932 931 930 929 928 927 926 925 924 923 922 921 920 919 918 917 916 915 914 913 912 911 910 909 908 907 906 905 904 903 902 901 900 899 898 897 896 895 894 893 892 891 890 889 888 887 886 885 884 883 882 881 880 879 878 877 876 875 874 873 872 871 870 869 868 867 866 865 864 863 862 861 860 859 858 857 856 855 854 853 852 851 850 849 848 847 846 845 844 843 842 841 840 839 838 837 836 835 834 833 832 831 830 829 828 827 826 825 824 823 822 821 820 819 818 817 816 815 814 813 812 811 810 809 808 807 806 805 804 803 802 801 800 799 798 797 796 795 794 793 792 791 790 789 788 787 786 785 784 783 782 781 780 779 778 777 776 775 774 773 772 771 770 769 768 767 766 765 764 763 762 761 760 759 758 757 756 755 754 753 752 751 750 749 748 747 746 745 744 743 742 741 740 739 738 737 736 735 734 733 732 731 730 729 728 727 726 725 724 723 722 721 720 719 718 717 716 715 714 713 712 711 710 709 708 707 706 705 704 703 702 701 700 699 698 697 696 695 694 693 692 691 690 690 689 688 687 686 685 684 683 682 681 680 680 679 678 677 676 675 674 673 672 671 670 670 669 668 667 666 665 664 663 662 661 660 660 659 658 657 656 655 654 653 652 651 650 650 649 648 647 646 645 644 643 642 641 640 640 639 638 637 636 635 634 633 632 631 630 630 629 628 627 626 625 624 623 622 621 620 620 619 618 617 616 615 614 613 612 611 610 610 609 608 607 606 605 604 603 602 601 600 600 599 598 597 596 595 594 593 592 591 590 590 589 588 587 586 585 584 583 582 581 580 580 579 578 577 576 575 574 573 572 571 570 570 569 568 567 566 565 564 563 562 561 560 560 559 558 557 556 555 554 553 552 551 550 550 549 548 547 546 545 544 543 542 541 540 540 539 538 537 536 535 534 533 532 531 530 530 529 528 527 526 525 524 523 522 521 520 520 519 518 517 516 515 514 513 512 511 510 510 509 508 507 506 505 504 503 502 501 500 500 499 599 498 497 496 495 494 493 492 491 490 490 489 488 487 486 485 484 483 482 481 480 480 479 478 477 476 475 474 473 472 471 470 470 469 468 467 466 465 464 463 462 461 460 460 459 458 457 456 455 454 453 452 451 450 450 449 448 447 446 445 444 443 442 441 440 440 439 438 437 436 435 434 433 432 431 430 430 429 428 427 426 425 424 423 422 421 420 420 419 418 417 416 415 414 413 412 411 410 410 409 408 407 406 405 404 403 402 401 400 400 399 398 397 396 395 394 393 392 391 390 390 389 388 387 386 385 384 383 382 381 380 380 379 378 377 376 375 374 373 372 371 370 370 369 368 367 366 365 364 363 362 361 360 360 359 358 357 356 355 354 353 352 351 350 350 349 348 347 346 345 344 343 342 341 340 340 339 338 337 336 335 334 333 332 331 330 330 329 328 327 326 325 324 323 322 321 320 320 319 318 317 316 315 314 313 312 311 310 310 309 308 307 306 305 304 303 302 301 300 300 299 298 297 296 295 294 293 292 291 290 290 289 288 287 286 285 284 283 282 281 280 280 279 278 277 276 275 274 273 272 271 270 270 269 268 267 266 265 264 263 262 261 260 260 259 258 257 256 255 254 253 252 251 250 250 249 248 247 246 245 244 243 242 241 240 240 239 238 237 236 235 234 233 232 231 230 230 229 228 227 226 225 224 223 222 221 220 220 219 218 217 216 215 214 213 212 211 210 210 209 208 207 206 205 204 203 202 201 200 200 199 198 197 196 195 194 193 192 191 190 190 189 188 187 186 185 184 183 182 181 180 180 179 178 177 176 175 174 173 172 171 170 170 169 168 167 166 165 164 163 162 161 160 160 159 158 157 156 155 154 153 152 151 150 150 149 148 147 146 145 144 143 142 141 140 140 139 138 137 136 135 134 133 132 131 130 130 129 128 127 126 125 124 123 122 121 120 120 119 118 117 116 115 114 113 112 111 110 110 109 108 107 106 105 104 103 102 101 100 100 99 98 97 96 95 94 93 92 91 90 90 89 88 87 86 85 84 83 82 81 80 80 79 78 77 76 75 74 73 72 71 70 70 69 68 67 66 65 64 63 62 61 60 60 59 58 57 56 55 54 53 52 51 50 50 49 48 47 46 45 44 43 42 41 40 40 39 38 37 36 35 34 33 32 31 30 30 29 28 27 26 25 24 23 22 21 20 20 19 18 17 16 15 14 13 12 11 10 10 9 8 7 6 5 4 3 2 1 0

a. contacting a target nucleic acid with a first set of pools of immobilized probes and a first set of pools of labeled probes, wherein at least one pool in either the first set of pools of immobilized probes, or in the first set of pools of labeled probes, or in both, comprises a mixture of two or more probes having different sequences in the information regions of the probes, under conditions which produce, on average, more probe:target hybridization for probes which are perfectly complementary to the target nucleic acid in the information region than with probes which are mismatched in the information region;

b. covalently joining adjacently hybridized immobilized probes and labeled probes to provide a first set of covalently joined probes;

c. detecting a first subset of pools of covalently joined probes for which a level of hybridization indicates that there is at least one perfectly complementary covalently joined probe within each pool; and

d. identifying one or more sequences of the target nucleic acid from the first subset of covalently joined pools of probes detected in step (c) by compiling overlapping sequences of the information regions of covalently joined probes in the subset of detected pools, wherein one or more covalently joined pooling false positive probes are eliminated as a result of compilation of overlapping sequences.

21. The method of claim 20 further comprising, following step (c) and before step (d), the steps of:

a. contacting the target nucleic acid with a second set of pools of immobilized probes and a second set of pools of labeled probes, wherein at least one probe in the second set of immobilized probes has the same information region as a probe in the first set of pools of immobilized probes, or at least one probe in the second set of labeled probes has the same information region as a probe in the first set of pools of labeled probes,

b. covalently joining adjacently hybridized immobilized probes and labeled probes to provide a second set of covalently joined probes;

c. detecting a second subset of covalently joined pools of probes for which a level of hybridization indicates that there is at least one perfectly complementary probe within each pool; and

5        d.     eliminating covalently joined probes with the same information regions present in both the first set of covalently joined pools of probes and the second set of covalently joined pools of probes that are not present in both the first detected subset of covalently joined pools of probes and the second detected subset of covalently joined pools of probes.

22.    A method of identifying one or more sequences of a target nucleic acid comprising:

10      a.     contacting a target nucleic acid with a first set of pools of immobilized probes and a first set of pools of labeled probes, wherein at least one pool in either the first set of pools of immobilized probes, or in the first set of pools of labeled probes, or in both, comprises a mixture of two or more probes having different sequences in the information regions of the probes, under conditions which produce, on average, more probe:target hybridization for probes which are perfectly complementary to the target nucleic acid in the information region than with probes which are mismatched in the information region;

15      b.     covalently joining adjacently hybridized immobilized probes and labeled probes to provide a first set of covalently joined probes;

20      c.     assigning a hybridization score to each covalently joined probe in the first set wherein each probe within a pool of covalently joined probes is assigned the same hybridization score, and

25      e.     identifying one or more sequences of the target nucleic acid from overlapping covalently joined probes by analysis of hybridization scores of overlapping covalently joined probes wherein one or more covalently joined probes with false high scores arising from pooling of probes are eliminated by analysis of hybridization scores of overlapping probes.

30      23.    The method of claim 22 further comprising after step (c) and before step (d) the steps of:

a.     contacting the target nucleic acid with a second set of pools of immobilized probes and a second set of pools of labeled probes, wherein at least one probe in the second set of immobilized probes has the same information region as a probe in the first set of pools of immobilized probes, or at least one probe in the second set of

D522626 D522626 D522626 D522626 D522626

labeled probes has the same information region as a probe in the first set of pools of labeled probes,

b. covalently joining adjacently hybridized immobilized probes and labeled probes to provide a second set of covalently joined probes;

5 c. assigning a hybridization score to each covalently joined probe of the second set wherein each probe within a pool of covalently joined probes is assigned the same hybridization score.

10 24. The method of claim 23 further comprising the step of  
d. eliminating the higher of two scores for covalently joined probes present in both the first set and second set of covalently joined pools of probes.

15 25. The method of claim 21, 23 or 24 wherein the first and second sets of pools of immobilized probes, or the first and second sets of pools of labeled probes, or both, comprise the same information regions.

20 26. The method of claim 21, 23 or 24 wherein the first and second sets of pools of immobilized probes, or the first and second sets of pools of labeled probes, or both, comprise the same probes.

27. The method of any one of claims 20 through 24 in which a label of the labeled probe is a fluorophore.

25 28. The method of any one of claims 20 through 24 in which a label of the labeled probe is attached to a terminal nucleotide.

29. The method of any one of claims 20 through 24 in which a label of the labeled probe is attached to an internal nucleotide.

30 30. The method of any one of claims 20 through 24 in which the set of pools of immobilized probes is immobilized on one or more solid supports.

31. The method of claim 30 in which the sets of pools of immobilized probes are arranged in a spatially-addressable array in which each pool has a unique address.

5 32. The method of claim 22, 23 or 24 wherein a statistical analysis of hybridization scores is performed.

33. The method of claim 22 wherein step (d) further comprises calculating a score for the identified one or more sequences of the target nucleic acid.

10 34. The method of any one of claims 20 through 24 wherein the pools of immobilized probes each consist of one probe.

15 35. The method of any one of claims 20 through 24 wherein the pools of labeled probes each consist of one probe.

20 36. A set of pools of probes wherein each probe comprises an information region, wherein said set of probes is sufficient to determine the sequence of an unknown target nucleic acid by overlapping sequences of the information region of two or more of said probes, and wherein at least one pool comprises two or more probes having different sequences in the information regions and having the same label or no label, and wherein the set of the pools of probes also satisfies one or more of the following rules describing the information regions of the probes, said rules selected from the group consisting of:

25           a. a consensus sequence of at least one pool in the set consists only of the letters selected from the group consisting of V, H, D, B, and N;

              b. a consensus sequence of probes in each pool in the set comprises more than three different letters selected from the group consisting of A, C, G, T, U, M, R, W, S, Y, K, V, H, D, B, and N;

30           c. consensus sequences from each informative position of all pools in the set comprise more than eight letters selected from the group

DOCUMENT EVIDENCE

consisting of A, C, G, T, U, M, R, W, S, Y, K, V, H, D, B, and N; and

5           d. consensus sequences from each information region of all pools in the set comprise more than five different letters selected from the group consisting of A, C, G, T, U, M, R, W, S, Y, K, V, H, D, B, and N, wherein at least one letter is selected from the group consisting of M, R, W, S, Y, and K.

10           37. The set of pools of probes of claim 36 wherein said set comprises all possible probes of the same length K, where K is greater than 3.

15           38. The set of pools of probes of claim 36 wherein each pool comprises more than 16 different probes.

20           39. The set of pools of probes of claim 38 wherein each pool comprises at least 32 different probes.

40. The set of pools of probes of claim 36 in which the pools are arranged in a spatially-addressable array, and wherein each pool has an address.

20           41. The set of pools of probes of claim 36 wherein at least two pools are mixed, wherein any two pools that are mixed are associated with different labels, and wherein all probes in a single pool are associated with the same label.

SEARCHED  
INDEXED  
SERIALIZED  
FILED